# Protocole Data

Table des matières

[Protocole Data 1](#_Toc13823988)

[Présentation 2](#_Toc13823989)

[Projet 2](#_Toc13823990)

[Objectifs 2](#_Toc13823991)

[Contenu 2](#_Toc13823992)

[Réflexion sur la thématique 2](#_Toc13823993)

[Principe ethique Data-scientist 2](#_Toc13823994)

[Récupération de la donnée 3](#_Toc13823995)

[Principe ethique Data-scientist 3](#_Toc13823996)

[Données recueillies 3](#_Toc13823997)

[Données Utilisateur 4](#_Toc13823998)

[Stockage de la donnée 4](#_Toc13823999)

[Traitements 4](#_Toc13824000)

[Principe ethique Data-scientist 4](#_Toc13824001)

[Traitements fonctionnels 5](#_Toc13824002)

[Traitements Métiers 5](#_Toc13824003)

[Analyse 5](#_Toc13824004)

[Principe ethique Data-scientist 5](#_Toc13824005)

[Analyse des résultats 5](#_Toc13824006)

[Prédictive 5](#_Toc13824007)

[Diffusion 6](#_Toc13824008)

[Principe ethique Data-scientist 6](#_Toc13824009)

[Datavisualisation 6](#_Toc13824010)

[Mise à disposition En OpenData 6](#_Toc13824011)

[Principe ethique Data-scientist 6](#_Toc13824012)

[Réutilisation 7](#_Toc13824013)

[Méthodologie de travail 7](#_Toc13824014)

## Présentation

Ce word va servir pour noter les étapes techniques nécessaires au traitement de la donnée.

Une sorte de protocole inspiré de la méthode scientifique.

### Projet

Création d’une web app (multi plateforme et responsive) permettant de recueillir l’opinion des 15-25 ans et jeunes adultes face à des situations ou comportements pouvant engendrer ou résulter de violences sexistes et sexuelles.

### Objectifs

Construire une image de la perception des 15-25 ans et jeunes adultes sur leur acceptation moral ou non de comportements pouvant engendrer et/ou étant des violences sexistes et sexuelles.

### Contenu

#### Front-End :

● Un questionnaire d’une série de 12 situations où le participant, en tant qu’observateur, devra juger s’il la pense acceptable ou inacceptable.

● En fin de questionnaire, une restitution de la position de ses réponses par rapport aux réponses moyennes générales.

#### Back-End

● Création d’une base de donnée dynamique

● Générer l’export des données ​sous forme de ​représentations graphiques et rapports statistiques

Il y a en parallèle la mise en place du projet, la gestion d’équipe et la gestion technique.

## Réflexion sur la thématique

### Principe ethique Data-scientist

« Je participe à un projet orienté "données"

Je m'engage à :

* Interroger la **finalité** du projet, sa **légalité** et son possible **impact social et environnemental**.
* Faire mon possible pour que mes conditions de travail, en particulier le **temps de travail** qui m’est alloué, me permettent de mener avec les données un travail honnête et **le plus scientifique possible**.
* Veiller à ce que les **métriques à optimiser** soient pertinentes et **ne conduisent pas** le projet à avoir un impact social et environnemental négatif. »

## Récupération de la donnée

Les données peuvent être rangées en 2 catégories :

* Données recueillies via sondage
* Données Utilisateur

### Principe ethique Data-scientist

« Je collecte ou je dispose de données

Je m'engage à :

* **Communiquer**, ou sinon rappeler aux équipes compétentes, la nécessité de communiquer auprès des personnes concernées, l’usage qui sera fait de leurs données, de **la façon la plus claire, explicite et transparente possible**.
* Veiller à ce que le **consentement des personnes** dont je collecte les données soit obtenu dans des conditions loyales et transparentes pour eux. En cas de changements ultérieurs des conditions d'utilisation, veiller à ce que ces changements leur soient aussi communiqués clairement et efficacement, et ré-obtenir leur consentement dans des conditions loyales, explicites et transparentes pour eux.
* Veiller à ce que les données dont j’ai la responsabilité soient **gérées et stockées en sécurité**.
* Ne pas collecter ou utiliser de **données inutilement personnelles et/ou sensibles**.
* Ne pas collecter **des données dont je peux raisonnablement estimer qu’elles ne seront pas utiles** par rapport aux besoins du projet.
* Ce que les données que je collecte ou dont je dispose soient **exactes** et que je comprenne bien leur signification. Pour cela, je m’engage à retracer autant que possible **l'origine** et le processus de création des données, ainsi que les éventuelles **modifications** qu'elles ont subies.
* Ne pas **négliger des données potentiellement utiles au projet**, dans la limite des principes éthiques, afin de ne pas mettre en péril la robustesse et la pertinence des résultats de leur traitement. »

### Données recueillies

Les données recueillies via l’application permettent de faire une cartographie des valeurs d’une personne.

Il s’agit d’un instantané unique car chaque scène sera unique par la variation des variables affichées et le profil de l’utilisateur.

Grâce à l’ensemble des résultats cela permettra de faire une moyenne des valeurs.

Moyenne qui devrait refléter l’état actuel de la société.

### Données Utilisateur

Les données utilisateur sont les données les plus importantes car elles permettent de coller un contexte aux réponses du questionnaire.

On peut considérer que les informations utilisateurs sont les informations les plus intéressantes car cela va permettre de :

* Juger la fréquentation du site
* Faire une cartographie des valeurs en fonction du milieu social et du sexe
* Faire des remontées d’expérience de vie

## Stockage de la donnée

La base de données conseillée est Postgresql pour son respect des différentes normes et sa scalabilité.

## Traitements

### Principe ethique Data-scientist

« Je prépare et j’explore les données

Je m'engage à :

* Ne pas créer de données caractéristiques (“features”) qui équivaudraient à des données personnelles sensibles si leur usage peut entraîner des effets discriminatoires illégaux ou illégitimes (exemples : code postal, nom de famille, ...).
* Veiller, lorsque j'élimine ou j'impute des valeurs manquantes ou aberrantes, à **ne pas introduire de biais supplémentaires** qui mèneraient à des résultats partiels ou faux. Pour cela :
* Je regarde la **distribution** des données à ma disposition ;
* Je m'interroge sur **leurs potentiels biais**, notamment le biais de sélection ;
* Je **justifie et documente** mon nettoyage. »

### Traitements fonctionnels

Typage

Meta données

Encodage

### Traitements Métiers

Faire des indicateurs spécifiques répondant aux besoins et permettant de comprendre les situations. Il serait préférable de faire appel à un un spécialiste de la thématique pour faire ressortir les informations essentielles.

## Analyse

### Principe ethique Data-scientist

« J'applique un ou des modèle(s) algorithmique(s)

Je m'engage à :

* Veiller à ce que le responsable du système puisse autant que possible **fournir une explication des résultats du modèle algorithmique** aux personnes concernées et ce d’autant plus s’il est légalement tenu d’expliquer ces décisions.
* Mesurer le **biais** et la **variance** pour contrôler l'exactitude et la dispersion du résultat et documenter les **métriques d'erreur** retenues.
* Déterminer le **meilleur compromis entre la performance et l'interprétabilité** sur l’ensemble des modèles à disposition et autant que possible, opter pour **les modèles les plus simples à expliquer aux personnes concernées** (un modèle performant permettra de diminuer les risques d’erreur tandis qu’un modèle interprétable permettra de mieux justifier les résultats du modèle).
* Prévoir et prévenir les **dérives possibles dans le temps du modèle** par rapport aux données, de façon à éviter l'apparition de biais supplémentaires.
* Paramétrer et **tester plusieurs modèles** en ne m’arrêtant pas au premier modèle et paramétrage qui me semblent bons. https://hippocrate.s3.eu-west-3.amazonaws.com/46313376580b4c44b6b1336de368dc52.pnghttps://hippocrate.s3.eu-west-3.amazonaws.com/46313376580b4c44b6b1336de368dc52.png
* Garder un esprit critique par rapport aux **segments** issus d'un **algorithme de groupement** (“clustering”). »

### Analyse des résultats

L’analyse des résultats se fera en croisant les données utilisateurs et les donnés recueillies par l’application.

### Prédictive

L’équipe n’avait pas cette compétence en interne.

## Diffusion

### Principe ethique Data-scientist

« Je présente les résultats aux parties prenantes

Je m'engage à :

* Donner **l'alerte** si je constate une utilisation frauduleuse, illégale, illégitime, discriminatoire ou non-éthique des résultats.
* Communiquer ma démarche et mes résultats (ou leur absence) à mon client / mon équipe :
* en ne les **dénaturant** et en ne les **dissimulant pas**, notamment en choisissant une visualisation fidèle à l’ensemble des résultats
* en **garantissant la compréhension la plus exacte possible**, en optant pour les visualisations et explications les plus parlantes, en précisant les **précautions d’usage** à prendre avec ces résultats et leur interprétation. »

### Datavisualisation

La datavisualistion doit être simple au départ afin de comprendre la donnée et l’interpréter.

Lorsque ce travail sera fait il faudra faire appel à un graphiste afin de transformer la donnée en information facilement compréhensible par tous.

## Mise à disposition En OpenData

### Principe ethique Data-scientist

« Je termine le projet

Je m'engage à :

* Anticiper les **usages** qui pourraient être faits de mon travail à moyen et long terme et faire mon possible pour garder sur lui un **droit de regard** et une **possibilité d’action**.
* Veiller à ce que les données ne soient pas conservées **plus longtemps que ce qui est nécessaire** pour l’application telle que définie pendant le projet.
* **Documenter au maximum les données et leurs traitements** afin d’en garantir l’**explicabilité** et la **reproductibilité**. »

Vu le contexte de l’observatoire sur les violences faites aux femmes, on peut considérer que les données sont d’intérêts publics.

En conséquence il serait intéressant de publier les données sur un site tel que <https://www.data.gouv.fr/fr/>

Ce qui permettrait une diffusion plus large auprès du public.

### Réutilisation

La mise en OpenData va permettre une centralisation des réutilisations de la part des organismes/associations/citoyens.

Ces réutilisations si croisées avec d’autres données externes peuvent apporter une grande valeur ajoutée.

# Méthodologie de travail

A Chaque étape il faut répondre à ces questions simples :

|  |  |
| --- | --- |
| Définition |  |
| Qui ? |  |
| Quoi ? |  |
| Où ? |  |
| Quand ? |  |
| Comment ? |  |
| Dans quel But ? |  |
| Législation particulière ? |  |